Sparse Recovery from Combined Fusion Frame Measurements

Petros Boufounos, *Member, IEEE,* Gitta Kutyniok, and Holger Rauhut

***Abstract*—Sparse representations have emerged as a powerful tool in signal and information processing, culminated by the success of new acquisition and processing techniques such as Compressed Sensing (CS). Fusion frames are very rich new signal representation methods that use collections of subspaces instead of vectors to represent signals. This work combines these exciting fields to introduce a new sparsity model for fusion frames. Signals that are sparse under the new model can be compressively sampled and uniquely reconstructed in ways similar to sparse signals using standard CS. The combination provides a promising new set of mathematical tools and signal models useful in a variety of applications. With the new model, a sparse signal has energy in very few of the subspaces of the fusion frame, although it does not need to be sparse within each of the subspaces it occupies. This sparsity model is captured using a mixed ℓ_1/ℓ_2 norm for fusion frames.**

A signal sparse in a fusion frame can be sampled using very few random projections and exactly reconstructed using a convex optimization that minimizes this mixed ℓ_1/ℓ_2 norm. The provided sampling conditions generalize coherence and RIP conditions used in standard CS theory. It is demonstrated that they are sufficient to guarantee sparse recovery of any signal sparse in our model. Moreover, a probabilistic analysis is provided using a stochastic model on the sparse signal that shows that under very mild conditions the probability of recovery failure decays exponentially with increasing dimension of the subspaces.

***Index Terms*—Compressed sensing, ℓ_1 minimization, $\ell_{1,2}$-minimization, sparse recovery, mutual coherence, fusion frames, random matrices.**

I. Introduction

COMPRESSED Sensing (CS) has recently emerged as a very powerful field in signal processing, enabling the acquisition of signals at rates much lower than previously thought possible [1], [2]. To achieve such performance, CS exploits the structure inherent in many naturally occurring and man-made signals. Specifically, CS uses classical signal representations and imposes a sparsity model on the signal of interest. The sparsity model, combined with randomized linear acquisition, guarantees that non-linear reconstruction can be used to efficiently and accurately recover the signal.

Fusion frames are recently emerged mathematical structures that can better capture the richness of natural and man-made signals compared to classically used representations [3]. In particular, fusion frames generalize frame theory by using subspaces in the place of vectors as signal building blocks. Thus signals can be represented as linear combinations of components that lie in particular, and often overlapping, signal subspaces. Such a representation provides significant flexibility in representing signals of interest compared to classical frame representations.

In this paper we extend the concepts and methods of Compressed Sensing to fusion frames. In doing so we demonstrate that it is possible to recover signals from underdetermined measurements if the signals lie only in very few subspaces of the fusion frame. Our generalized model does not require that the signals are sparse within each subspace. The rich structure of the fusion frame framework allows us to characterize more complicated signal models than the standard sparse or compressible signals used in compressed sensing techniques. This paper complements and extends our work in [4].

Introducing sparsity in the rich fusion frame model and introducing fusion frame models to the Compressed Sensing literature is a major contribution of this paper. We extend the results of the standard worst-case analysis frameworks in Compressed Sensing, using the sampling matrix *null space property (NSP)*, *coherence*, and *restricted isometry property (RIP)*. In doing so, we extend the definitions to fusion NSP, fusion coherence and fusion RIP to take into account the differences of the fusion frame model. The three approaches provide complementary intuition on the differences between standard sparsity and block-sparsity models and sparsity in fusion frame models. We note that in the special case that the subspaces in our model all have the same dimension, most of our results follow from previous analysis of block-sparsity models [5], [6]. But in the general case of different dimensions they are new.

Our understanding of the problem is further enhanced by the probabilistic analysis. As we move from standard sparsity to fusion frame or other vector-based sparsity models, worst case analysis becomes increasingly pessimistic. The probabilistic analysis provides a framework to discern which assumptions of the worst case model become irrelevant and which are

Manuscript received December 10, 2009; revised November 16, 2010; accepted December 14, 2010. Date of current version May 25, 2011. P. Boufounos did part of this work while at Rice university. P. Boufounos was supported in part by the grants NSF CCF-0431150, CCF-0728867, CNS-0435425, and CNS-0520280, DARPA/ONR N66001-08-1-2065, ONR N00014-07-1-0936, N00014-08-1-1067, N00014-08-1-1112, and N00014-08-1-1066, AFOSR FA9550-07-1-0301, ARO MURI W311NF-07-1-0185, and the Texas Instruments Leadership University Program. P. Boufounos was supported in part by Mitsubishi Electric Research Laboratories (MERL). G. Kutyniok was in part supported by Deutsche Forschungsgemeinschaft (DFG) Heisenberg Fellowship KU 1446/8, DFG Grant SPP-1324, KU 1446/13, and DFG Grant KU 1446/14. H. Rauhut acknowledges support by the Hausdorff Center for Mathematics and by the WWTF project SPORTS (MA 07-004).

P. Boufounos is with Mitsubishi Electric Research Laboratories, *petrosb@merl.com*.

G. Kutyniok is with Institute of Mathematics, University of Osnabrück, *kutyniok@uni-osnabrueck.de*.

H. Rauhut is with Hausdorff Center for Mathematics & Institute for Numerical Simulation, University of Bonn, *rauhut@hcm.uni-bonn.de*.

Communicated by J. Romberg, Associate Editor for Signal Processing.

Digital Object Identifier 10.1109/TIT.2011.2143890

critical. It further demonstrates the significance of the angles between the subspaces comprising the fusion frame. Although our analysis is inspired by the model and the analysis in [7], the tools used in that work do not extend to the fusion frame model. The analysis presented in Sec. V is the second major contribution of our paper.

In the remainder of this section we provide the motivation behind our work and describe some possible applications. Section II provides some background on Compressed Sensing and on fusion frames to serve as a quick reference for the fundamental concepts and our basic notation. In Section III we formulate the problem, establish the additional notation and definitions necessary in our development, and state the main results of our paper. We further explore the connections with existing research in the field, as well as possible extensions. In Section IV we prove deterministic recovery guarantees using the properties of the sampling matrix. Section V presents the probabilistic analysis of our model, which is more appropriate for typical usage scenarios. We conclude with a discussion of our results.

A. Motivation

As technology progresses, signals and computational sensing equipment becomes increasingly multidimensional. Sensors are being replaced by sensor arrays and samples are being replaced by multidimensional measurements. Yet, modern signal acquisition theory has not fully embraced the new computational sensing paradigm. Multidimensional measurements are often treated as collections of one-dimensional ones due to the mathematical simplicity of such treatment. This approach ignores the potential information and structure embedded in multidimensional signal and measurement models.

Our ultimate motivation is to provide a better understanding of more general mathematical objects, such as vector-valued data points [8]. Generalizing the notion of sparsity is part of such understanding. Towards that goal, we demonstrate that the generalization we present in this paper encompasses joint sparsity models [9], [10] as a special case. Furthermore, it is itself a special case of block-sparsity models [11], [12], [5], [6], with significant additional structure.

B. Applications

Although the development in this paper provides a general theoretical perspective, the principles and the methods we develop are widely applicable. In particular, the special case of joint (or simultaneous) sparsity has already been widely used in radar [13], sensor arrays [14], and MRI pulse design [15]. In these applications a mixed ℓ_1/ℓ_2 norm was used heuristically as a sparsity proxy. Part of our goals in this paper is to provide a solid theoretical understanding of such methods.

In addition, the richness of fusion frames allows the application of this work to other cases, such as target recognition and music segmentation. The goal in such applications is to identify, measure, and track targets that are not well described by a single vector but by a whole subspace. In music segmentation, for example, each note is not characterized by a single frequency, but by the subspace spanned by the fundamental frequency of the instrument and its harmonics [16]. Furthermore, depending on the type of instrument in use, certain harmonics might or might not be present in the subspace. Similarly, in vehicle tracking and identification, the subspace of a vehicle's acoustic signature depends on the type of vehicle, its engine and its tires [17]. Note that in both applications, there might be some overlap in the subspaces which distinct instruments or vehicles occupy.

Fusion frames are quite suitable for such representations. The subspaces defined by each note and each instrument or each tracked vehicle generate a fusion frame for the whole space. Thus the fusion frame serves as a dictionary of targets to be acquired, tracked, and identified. The fusion frame structure further enables the use of sensor arrays to perform joint source identification and localization using far fewer measurements than a classical sampling framework. In Sec. III-E we provide a stylized example that demonstrates this potential.

We also envision fusion frames to play a key role in video acquisition, reconstruction and compression applications such as [18]. Nearby pixels in a video exhibit similar sparsity structure locally, but not globally. A joint sparsity model such as [9], [10] is very constraining in such cases. On the other hand, subspace-based models for different parts of an image significantly improve the modeling ability compared to the standard compressed sensing model.

C. Notation

Throughout this paper $\|\mathbf{x}\|_p = \left(\sum_i x_i^p\right)^{1/p},\ p > 0$ denotes the standard ℓ_p norm. The operator norm of a matrix A from ℓ_p into ℓ_p is written as $\|A\|_{p\to p} = \max_{\|x\|_p \le 1} \|Ax\|_p$.

II. Background

A. Compressed Sensing

Compressed Sensing (CS) is a recently emerged field in signal processing that enables signal acquisition using very few measurements compared to the signal dimension, as long as the signal is sparse in some basis. It predicts that a signal $\mathbf{x} \in \mathbb{R}^N$ with only k non-zero coefficients can be recovered from only $n = \mathcal{O}(k \log(N/k))$ suitably chosen linear non-adaptive measurements, compactly represented by

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad \mathbf{y} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{n\times N}.$$

A necessary condition for exact signal recovery of all k-sparse $\mathbf{x}$ is that

$$\mathbf{A}\mathbf{z} \neq 0 \quad \text{ for all } \mathbf{z} \neq 0, \|\mathbf{z}\|_0 \le 2k,$$

where the ℓ_0 'norm,' $\|\mathbf{x}\|_0$, counts the number of non-zero coefficients in $\mathbf{x}$. In this case, recovery is possible using the following combinatorial optimization,

$$\widehat{\mathbf{x}} = \text{argmin}_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}.$$

Unfortunately this is an NP-hard problem [19] in general, hence is infeasible.

Exact signal recovery using computationally tractable methods can be guaranteed if the measurement matrix $\mathbf{A}$ satisfies

the null space property (NSP) [20], [21], i.e., if for all support sets $S \subset \{1, \ldots, N\}$ of cardinality at most k,

$$\|\mathbf{h}_S\|_1 < \frac{1}{2}\|\mathbf{h}\|_1 \quad \text{for all } \mathbf{h} \in \mathrm{nullsp}(\mathbf{A}) \setminus \{0\},$$

where $\mathbf{h}_S$ denotes the vector which coincides with $\mathbf{h}$ on the index set S and is zero outside S.

If the coherence of $\mathbf{A}$ is sufficiently small, the measurement matrix satisfies the NSP and, therefore, exact recovery is guaranteed [22], [8]. The *coherence* of a matrix $\mathbf{A}$ with unit norm columns $\mathbf{a}_i$, $\|\mathbf{a}_i\|_2 = 1$, is defined as

$$\mu = \max_{i \neq j} |\langle \mathbf{a}_i, \mathbf{a}_j \rangle| \,. \tag{1}$$

Exact signal recovery is also guaranteed if $\mathbf{A}$ obeys a *restricted isometry property (RIP) of order* $2k$ [1], i.e., if there exists a constant $\delta_{2k} \in (0, 1)$ such that for all $2k$-sparse signals $\mathbf{x}$

$$(1 - \delta_{2k})\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_{2k})\|\mathbf{x}\|_2^2.$$

We note the relation $\delta_{2k} \leq (k-1)\mu$, which easily follows from Gershgorin's theorem. If any of the above properties hold, then the following convex optimization program exactly recovers the signal from the measurement vector $\mathbf{y}$,

$$\widehat{\mathbf{x}} = \mathrm{argmin}_{\,\mathbf{x} \in \mathbb{R}^N} \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}.$$

A surprising result is that random matrices with sufficient number of rows can achieve small coherence and small RIP constants with overwhelmingly high probability.

A large body of literature extends these results to measurements of signals in the presence of noise, to signals that are not exactly sparse but compressible [1], to several types of measurement matrices [23], [24], [25], [26], [27] and to measurement models beyond simple sparsity [28].

B. Fusion Frames

Fusion frames are generalizations of frames that provide a richer description of signal spaces. A *fusion frame* for $\mathbb{R}^M$ is a collection of subspaces $\mathcal{W}_j \subseteq \mathbb{R}^M$ and associated weights v_j, compactly denoted by $(\mathcal{W}_j, v_j)_{j=1}^N$, that satisfies

$$A\|\mathbf{x}\|_2^2 \leq \sum_{j=1}^N v_j^2 \, \|\mathbf{P}_j \mathbf{x}\|_2^2 \leq B\|\mathbf{x}\|_2^2$$

for some universal fusion frame bounds $0 < A \leq B < \infty$ and for all $\mathbf{x} \in \mathbb{R}^M$, where $\mathbf{P}_j$ denotes the orthogonal projection onto the subspace $\mathcal{W}_j$. We use m_j to denote the dimension of the jth subspace $\mathcal{W}_j$, $j = 1, \ldots, N$. A frame is a special case of a fusion frame in which all the subspaces $\mathcal{W}_j$ are one-dimensional (i.e., $m_j = 1, \; j = 1, \ldots, N$), and the weights v_j are the norms of the frame vectors. For finite M and N, the definition reduces to the requirement that the subspace sum of $\mathcal{W}_j$ is equal to $\mathbb{R}^M$.

The generalization to fusion frames allows us to capture interactions between frame vectors to form specific subspaces that are not possible in classical frame theory. Similar to classical frame theory, we call the fusion frame *tight* if the frame bounds are equal, $A = B$. If the fusion frame has $v_j = 1, \; j = 1, \ldots, N$, we call it a *unit-norm* fusion frame. In this paper, we will in fact restrict to the situation of unit-norm fusion frames, since the anticipated applications are only concerned with membership in the subspaces and do not necessitate a particular weighting.

Dependent on a fusion frame $(\mathcal{W}_j, v_j)_{j=1}^N$ we define the Hilbert space $\mathcal{H}$ as

$$\begin{aligned} \mathcal{H} &= \{(\mathbf{x}_j)_{j=1}^N : \mathbf{x}_j \in \mathcal{W}_j \text{ for all } j = 1, \ldots, N\} \\ &\subseteq \mathbb{R}^{M \times N} (\text{or } \mathbb{R}^{MN}). \end{aligned}$$

We should point out that depending on the use, $\mathbf{x}$ can be represented as a very long vector or as a matrix. However, both representations are just rearrangements of vectors in the same Hilbert space, and we use them interchangeably in the manuscript.

Finally, let $\mathbf{U}_j \in \mathbb{R}^{M \times m_j}$ be a known but otherwise arbitrary matrix, the columns of which form an orthonormal basis for $\mathcal{W}_j$, $j = 1, \ldots, N$, that is $\mathbf{U}_j^T \mathbf{U}_j = \mathbf{I}_{m_j}$, where $\mathbf{I}_{m_j}$ is the $m_j \times m_j$ identity matrix, and $\mathbf{U}_j \mathbf{U}_j^T = \mathbf{P}_j$.

The *fusion frame mixed* $\ell_{q,p}$ *norm* is defined as

$$\left\|(\mathbf{x}_j)_{j=1}^N\right\|_{q,p} \equiv \left(\sum_{j=1}^N (v_j \|\mathbf{x}_j\|_q)^p \right)^{1/p}, \tag{2}$$

where $(v_j)_{j=1}^N$ are the fusion frame weights. Furthermore, for a sequence $\mathbf{c} = (\mathbf{c}_j)_{j=1}^N$, $\mathbf{c}_j \in \mathbb{R}^{m_j}$, we similarly define the mixed norm

$$\|\mathbf{c}\|_{2,1} = \sum_{j=1}^N \|\mathbf{c}_j\|_2.$$

The $\ell_{q,0}$–'norm' (which is actually not even a quasi-norm) is defined as

$$\|\mathbf{x}\|_{q,0} = \#\{j : \mathbf{x}_j \neq 0\},$$

independent of q. For the remainder of the paper we use $q = 2$ just for the purpose of notation and to distinguish $\ell_{2,0}$ from the ℓ_0 vector 'norm'. We call a vector $\mathbf{x} \in \mathcal{H}$ *k-sparse*, if $\|\mathbf{x}\|_{2,0} \leq k$.

III. Sparse Recovery of Fusion Frame Vectors

A. Measurement Model

We now consider the following scenario. Let $\mathbf{x}^0 = (\mathbf{x}_j^0)_{j=1}^N \in \mathcal{H}$, and assume that we only observe n linear combinations of those vectors, i.e., there exist some scalars a_{ij} satisfying $\|(a_{ij})_{i=1}^n\|_2 = 1$ for all $j = 1, \ldots, N$ such that we observe

$$\mathbf{y} = (\mathbf{y}_i)_{i=1}^n = \left(\sum_{j=1}^N a_{ij} \mathbf{x}_j^0 \right)_{i=1}^n \in \mathcal{K}, \tag{3}$$

where $\mathcal{K}$ denotes the Hilbert space

$$\mathcal{K} = \{(\mathbf{y}_i)_{i=1}^n : \mathbf{y}_i \in \mathbb{R}^M \text{ for all } i = 1, \ldots, n\}.$$

We first notice that (3) can be rewritten as

$$\mathbf{y} = \mathbf{A_I}\mathbf{x}^0, \quad \text{where } \mathbf{A_I} = (a_{ij}\mathbf{I}_M)_{1 \leq i \leq n,\, 1 \leq j \leq N},$$

i.e., $\mathbf{A_I}$ is the matrix consisting of the blocks $a_{ij}\mathbf{I}_M$.

B. Reconstruction using Convex Optimization

We now wish to recover $\mathbf{x}^0$ from the measurements $\mathbf{y}$. If we impose conditions on the sparsity of $\mathbf{x}^0$, it is suggestive to consider the following minimization problem,

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}\in\mathcal{H}} \|\mathbf{x}\|_{2,0} \quad \text{subject to } \sum_{j=1}^{N} a_{ij}\mathbf{x}_j = \mathbf{y}_i \text{ for all } i = 1,\ldots,n.$$

Using the matrix $\mathbf{A_I}$, we can rewrite this optimization problem as

$$(P_0) \quad \hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}\in\mathcal{H}} \|\mathbf{x}\|_{2,0} \quad \text{subject to } \mathbf{A_I}\mathbf{x} = \mathbf{y}.$$

However, this problem is NP-hard [19] and, as proposed in numerous publications initiated by [29], we prefer to employ ℓ_1 minimization techniques. This leads to the investigation of the following minimization problem,

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}\in\mathcal{H}} \|\mathbf{x}\|_{2,1} \quad \text{subject to } \mathbf{A_I}\mathbf{x} = \mathbf{y}.$$

Since we minimize over all $\mathbf{x} = (\mathbf{x}_j)_{j=1}^N \in \mathcal{H}$ and certainly $\mathbf{P}_j\mathbf{x}_j = \mathbf{x}_j$ by definition, we can rewrite this minimization problem as

$$(\tilde{P}_1) \quad \hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}\in\mathcal{H}} \|\mathbf{x}\|_{2,1} \quad \text{subject to } \mathbf{A_P}\mathbf{x} = \mathbf{y},$$

where

$$\mathbf{A_P} = (a_{ij}\mathbf{P}_j)_{1\le i\le n,\, 1\le j\le N}. \tag{4}$$

Problem $(\tilde{P}_1)$ bears difficulties to implement since minimization runs over $\mathcal{H}$. Still, it is easy to see that $(\tilde{P}_1)$ is equivalent to the optimization problem

$$(P_1) \quad (\hat{\mathbf{c}}_j)_j = \operatorname{argmin}_{\mathbf{c}_j\in\mathbb{R}^{m_j}} \|(\mathbf{U}_j\mathbf{c}_j)_{j=1}^N\|_{2,1} \quad \text{subject to } \mathbf{A_I}(\mathbf{U}_j\mathbf{c}_j)_j = \mathbf{y}, \tag{5}$$

where then $\hat{\mathbf{x}} = (\mathbf{U}_j\hat{\mathbf{c}}_j)_{j=1}^N$. This particular form ensures that the minimizer lies in the collection of subspaces $(\mathcal{W}_j)_{j=1}^N$ while minimization is performed over $\mathbf{c}_j \in \mathbb{R}^{m_j}$ for all $j = 1,\ldots,N$ and $\sum_j m_j \le MN$, hence feasible.

Finally, by rearranging (5), the optimization problems, invoking the ℓ_0-'norm' and ℓ_1-norm, can be rewritten using matrix-only notation as

$$(P_0) \quad \hat{\mathbf{c}} = \operatorname{argmin}_{\mathbf{c}} \|\mathbf{c}\|_{2,0} \text{ subject to } \mathbf{Y} = \mathbf{A}\mathbf{U}(\mathbf{c})$$

and

$$(P_1) \quad \hat{\mathbf{c}} = \operatorname{argmin}_{\mathbf{c}} \|\mathbf{c}\|_{2,1} \text{ subject to } \mathbf{Y} = \mathbf{A}\mathbf{U}(\mathbf{c}),$$

in which

$$\mathbf{U}(\mathbf{c}) = \begin{pmatrix} \mathbf{c}_1^T\mathbf{U}_1^T \\ \hline \vdots \\ \hline \mathbf{c}_N^T\mathbf{U}_N^T \end{pmatrix} \in \mathbb{R}^{N\times M},\ \mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 \\ \hline \vdots \\ \hline \mathbf{y}_n \end{pmatrix} \in \mathbb{R}^{n\times M},$$

$$\mathbf{A} = (a_{ij}) \in \mathbb{R}^{n\times N}, \quad \mathbf{c}_j \in \mathbb{R}^{m_j}, \text{ and } \mathbf{y}_i \in \mathbb{R}^M.$$

Hereby, we additionally used that $\|\mathbf{U}_j\mathbf{c}_j\|_2 = \|\mathbf{c}_j\|_2$ by orthonormality of the columns of $\mathbf{U}_j$. We follow this notation for the remainder of the paper.

C. Worst Case Recovery Conditions

To guarantee that $(\tilde{P}_1)$ always recovers the original signal $\mathbf{x}^0$, we provide three alternative conditions, in line with the standard CS literature [8], [23], [1], [20], [30], [31], [32], [22]. Specifically, we define the fusion null space property, the fusion coherence, and the fusion restricted isometry propery (FRIP). These are fusion frame versions of the null space property, the coherence and the RIP, respectively.

Definition 3.1: The pair $(\mathbf{A}, (\mathcal{W}_j)_{j=1}^N)$, with a matrix $\mathbf{A} \in \mathbb{R}^{n\times N}$ and a fusion frame $(\mathcal{W}_j)_{j=1}^N$ is said to satisfy the *fusion null space property* if

$$\|\mathbf{h}_S\|_{2,1} < \frac{1}{2}\|\mathbf{h}\|_{2,1} \quad \text{for all } \mathbf{h} \in \mathcal{N}\setminus\{0\},$$

for all support sets $S \subset \{1,\ldots,N\}$ of cardinality at most k. Here $\mathcal{N}$ denotes the null space $\{\mathbf{h} = (\mathbf{h}_j)_{j=1}^N : \mathbf{h}_j \in \mathbb{R}^{m_j}, \mathbf{A}\mathbf{U}(\mathbf{h}) = 0\}$, and $\mathbf{h}_S$ denotes the vector which coincides with $\mathbf{h}$ on the index set S and is zero outside S.

Definition 3.2: The *fusion coherence* of a matrix $\mathbf{A} \in \mathbb{R}^{n\times N}$ with normalized 'columns' $(\mathbf{a}_j = \mathbf{a}_{\cdot,j})_{j=1}^N$ and a fusion frame $(\mathcal{W}_j)_{j=1}^N$ for $\mathbb{R}^M$ is given by

$$\mu_f = \mu_f(\mathbf{A}, (\mathcal{W}_j)_{j=1}^N) = \max_{j\neq k}\left[|\langle \mathbf{a}_j, \mathbf{a}_k\rangle| \cdot \|\mathbf{P}_j\mathbf{P}_k\|_{2\to 2}\right],$$

where $\mathbf{P}_j$ denotes the orthogonal projection onto $\mathcal{W}_j$, $j = 1,\ldots,N$.

Definition 3.3: Let $\mathbf{A} \in \mathbb{R}^{n\times N}$ and $(\mathcal{W}_j)_{j=1}^N$ be a fusion frame for $\mathbb{R}^M$ and $\mathbf{A_P}$ as defined in (4). The *fusion restricted isometry constant* δ_k is the smallest constant such that

$$(1-\delta_k)\|\mathbf{z}\|_{2,2}^2 \le \|\mathbf{A_P}\mathbf{z}\|_{2,2}^2 \le (1+\delta_k)\|\mathbf{z}\|_{2,2}^2$$

for all $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N) \in \mathbb{R}^{MN}$, $\mathbf{z}_j \in \mathbb{R}^M$, of sparsity $\|\mathbf{z}\|_0 \le k$.

In the case that the subspaces have the same dimension the definitions of fusion coherence and fusion RIP coincide with those of block coherence and block RIP introduced in [5], [6]. This is expected, as we discuss in Sec. III-E, since the fusion sparsity model is a special case of general block sparsity models.

Using those definitions, we provide three alternative recovery conditions, also in line with standard CS results. We first state the characterization via the fusion null space property.

Theorem 3.4: Let $\mathbf{A} \in \mathbb{R}^{n\times N}$ and $(\mathcal{W}_j)_{j=1}^N$ be a fusion frame. Then all $\mathbf{c} = (\mathbf{c}_j)_{j=1}^N$, $\mathbf{c}_j \in \mathbb{R}^{m_j}$, with $\|\mathbf{c}\|_{2,0} \le k$ are the unique solution to (P_1) with $\mathbf{Y} = \mathbf{A}\mathbf{U}(\mathbf{c})$ if and only if $(\mathbf{A}, (\mathcal{W}_j)_{j=1}^N)$ satisfies the fusion null space property of order k.

While the fusion null space property characterizes recovery, it is somewhat difficult to check in practice. The fusion coherence is much more accessible for a direct calculation, and the next result states a corresponding sufficient condition. In the case, that the subspaces have the same dimension the theorem below reduces to Theorem 3 in [6] on block-sparse recovery.

Theorem 3.5: Let $\mathbf{A} \in \mathbb{R}^{n\times N}$ with normalized columns $(\mathbf{a}_j)_{j=1}^N$, let $(\mathcal{W}_j)_{j=1}^N$ be a fusion frame in $\mathbb{R}^M$, and let $\mathbf{Y} \in$

$\mathbb{R}^{n\times M}$. If there exists a solution $\mathbf{c}^0$ of the system $\mathbf{AU}(\mathbf{c}) = \mathbf{Y}$ satisfying

$$\|\mathbf{c}^0\|_{2,0} < \frac{1}{2}(1 + \mu_f^{-1}), \tag{6}$$

then this solution is the unique solution of (P_0) as well as of (P_1).

Finally, we state a sufficient condition based on the fusion RIP, which allows stronger recovery results, but is more difficult to evaluate than the fusion coherence. The constant $1/3$ below is not optimal and can certainly be improved, but our aim was rather to have a short proof. Furthermore, in the case that all the subspaces have the same dimension, the fusion frame setup is equivalent to the block-sparse case, for which the next theorem is implied by Theorem 1 in [5].

Theorem 3.6: Let $(\mathbf{A}, (\mathcal{W}_j)_{j=1}^N)$ with fusion frame restricted isometry constant $\delta_{2k} < 1/3$. Then (P_1) recovers all k-sparse $\mathbf{c}$ from $\mathbf{Y} = \mathbf{AU}(\mathbf{c})$.

D. Probability of Correct Recovery

Intuitively, it seems that the higher the dimension of the subspaces $\mathcal{W}_j$, the 'easier' the recovery via the ℓ_1 minimization problem [7] should be. However, it turns out that this intuition only holds true if we consider a probabilistic analysis. Thus we provide a probabilistic signal model and a typical case analysis for the special case where all the subspaces have the same dimension $m = m_j$ for all j. Inspired by the probability model in [7], we will assume that on the k-element support set $S = \text{supp}(\mathbf{U}(\mathbf{c})) = \{j_1, \ldots, j_k\}$ the entries of the vectors $\mathbf{c}_j$, $j \in S$, are independent and follow a normal distribution,

$$\mathbf{U}(\mathbf{c})_S = \begin{pmatrix} \mathbf{X}_1^T \mathbf{U}_{j_1}^T \\ \hline \vdots \\ \hline \mathbf{X}_k^T \mathbf{U}_{j_k}^T \end{pmatrix} \in \mathbb{R}^{k\times M}, \tag{7}$$

where $\mathbf{X} = (\mathbf{X}_1^T \ldots \mathbf{X}_k^T)^T \in \mathbb{R}^{Nm}$ is a Gaussian random vector, i.e., all entries are independent standard normal random variables.

Our probabilistic result shows that the failure probability for recovering $\mathbf{U}(\mathbf{c})$ decays exponentially fast with growing dimension m of the subspaces. Interestingly, the quantity θ involved in the estimate is again dependent on the 'angles' between subspaces and is of the flavor of the fusion coherence in Def. 3.2. Since the block-sparsity model can be seen as a special case of the fusion frame sparsity model, the theorem clearly applies also to this scenario.

Theorem 3.7: Let $S \subseteq \{1, \ldots, N\}$ be a set of cardinality k and suppose that $\mathbf{A} \in \mathbb{R}^{n\times N}$ satisfies

$$\|\mathbf{A}_S^\dagger \mathbf{A}_{\cdot,j}\|_2 \le \alpha < 1 \quad \text{for all } j \notin S. \tag{8}$$

Let $(\mathcal{W}_j)_{j=1}^N$ be a fusion frame with associated orthogonal bases $(\mathbf{U}_j)_{j=1}^N$ and orthogonal projections $(\mathbf{P}_j)_{j=1}^N$, and let $\mathbf{Y} \in \mathbb{R}^{n\times M}$. Further, let $\mathbf{c}_j \in \mathbb{R}^{m_j}$, $j = 1 \ldots, N$ with $S = \text{supp}(\mathbf{U}(\mathbf{c}))$ such that the coefficients on S are given by (7), and let θ be defined by

$$\theta = 1 + \max_{i\in S} \sum_{j\in S, j\neq i} \lambda_{\max}(\mathbf{P}_i\mathbf{P}_j)^{1/2}.$$

Choose $\delta \in (0, 1 - \alpha^2)$. Then with probability at least

$$1 - (N-k)\exp\left(-\frac{(\sqrt{1-\delta} - \alpha)^2}{2\alpha^2\theta} m\right) - k\exp\left(-\frac{\delta^2}{4} m\right)$$

the minimization problem (P_1) recovers $\mathbf{U}(\mathbf{c})$ from $\mathbf{Y} = \mathbf{AU}(\mathbf{c})$. In particular, the failure probability can be estimated by

$$N \exp\left(-\left(\max_{\delta\in(0,1-\alpha^2)} \min\left\{\frac{(\sqrt{1-\delta}-\alpha)^2}{2\alpha^2\theta}, \frac{\delta^2}{4}\right\}\right) m\right).$$

Let us note that [7, Section V] provides several mild conditions that imply (8). We exemplify one of these. Suppose that the columns of $A \in \mathbb{R}^{n\times N}$ form a unit norm tight frame with (ordinary) coherence $\mu \le c/\sqrt{n}$. (This condition is satisfied for a number of explicitly given matrices, see also [7].) Suppose further that the support set S is chosen uniformly at random among all subsets of cardinality k. Then with high probability (8) is satisfied provided $k \le C_\alpha n$, see Theorem 5.4 and Section V.A in [7]. This is in sharp contrast to deterministic recovery guarantees based on coherence, which cannot lead to better bounds than $k \le C\sqrt{n}$. Further note, that the quantity θ is bounded by k in the worst case, but maybe significantly smaller if the subspaces are close to orthogonal. Clearly, m can only be varied by changing the fusion frame model, and θ may also change in this case. While, in principle, θ may grow with m, typically θ actually decreases; for instance if the subspaces are chosen at random. In any case, since always $\theta \le k$, the failure probability of recovery decays exponentially in the dimension m of the subspaces.

E. Relation with Previous Work

A special case of the problem above appears when all subspaces $(\mathcal{W}_j)_{j=1}^N$ are equal and also equal to the ambient space $\mathcal{W}_j = \mathbb{R}^M$ for all j. Thus, $\mathbf{P}_j = \mathbf{I}_M$ and the observation setup of Eq. (3) is identical to the matrix product

$$\mathbf{Y} = \mathbf{AX}, \quad \text{where } \mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \hline \vdots \\ \hline \mathbf{x}_N \end{pmatrix} \in \mathbb{R}^{N\times M}. \tag{9}$$

This special case is the same as the well studied joint sparsity setup of [33], [9], [10], [34], [7] in which a collection of M sparse vectors in $\mathbb{R}^N$ is observed through the same measurement matrix $\mathbf{A}$, and the recovery assumes that all the vectors have the same sparsity structure. The use of mixed ℓ_1/ℓ_2 optimization has been proposed and widely used in this case.

The following simple example illustrates how a fusion frame model can reduce the sampling required compared to simple joint sparsity models. Consider the measurement scenario of (9), rewritten in an expanded form:

$$\begin{bmatrix} -\mathbf{y}_1^T- \\ \vdots \\ -\mathbf{y}_n^T- \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_N \\ | & | & & | \end{bmatrix} \begin{bmatrix} -\mathbf{x}_1^T- \\ -\mathbf{x}_2^T- \\ \vdots \\ -\mathbf{x}_N^T- \end{bmatrix},$$

where $\mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^M$ and $\mathbf{x}_i \in \mathcal{W}_i$. Using the fusion frames notation, $\mathbf{X}$ is a fusion frame representation of our acquired

signal, which we assume sparse as we defined in Sec. II-B. For the purposes of the example, suppose that only $\mathbf{x}_1$ and $\mathbf{x}_2$ have significant content and the remaining components are zero. Thus we only focus on the measurement vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ and their relationship, with respect to the subspaces $\mathcal{W}_1$ and $\mathcal{W}_2$ where $\mathbf{x}_1$ and $\mathbf{x}_2$ lie in.

Using the usual joint sparsity models, would require that $\mathbf{a}_1$ and $\mathbf{a}_2$ have low coherence (1), even if prior information about $\mathcal{W}_1$ and $\mathcal{W}_2$ provided a better signal model. For example, if we know from the problem formulation that the two components $\mathbf{x}_1$ and $\mathbf{x}_2$ lie in orthogonal subspaces $\mathcal{W}_1 \perp \mathcal{W}_2$, we can select $\mathbf{a}_1$ and $\mathbf{a}_2$ to be identical, and still be able to recover the signal. If, instead, $\mathcal{W}_1$ and $\mathcal{W}_2$ only have a common overlapping subspace we need $\mathbf{a}_1$ and $\mathbf{a}_2$ to be incoherent only when projected on that subspace, as measured by fusion coherence, defined in Def. 3.2, and irrespective of the dimensionality of the two subspaces and the dimensionality of their common subspace. This is a case not considered in the existing literature.

The practical applications are significant. For example, consider a wideband array signal acquisition system in which each of the signal subspaces are particular targets of interest from particular directions of interest (e.g. as an extremely simplified stylized example consider $\mathcal{W}_1$ as the subspace of friendly targets and $\mathcal{W}_2$ as the subspace of enemy targets from the same direction). If two kinds of targets occupy two different subspaces in the signal space, we can exploit this in the acquisition system. This opens the road to subspace-based detection and subspace-based target classification dictionaries.

Our formulation is a special case of the block sparsity problem [11], [12], [6], where we impose a particular structure on the measurement matrix $\mathbf{A}$. This relationship is already known for the joint sparsity model, which is also a special case of block sparsity. In other words, the fusion frame formulation we examine here specializes block sparsity problems and generalizes joint sparsity ones.

For the special case of fusion frames in which all the subspaces $\mathcal{W}_j$ have the same dimension, our definition of coherence can be shown to be essentially the same (within a constant scaling factor) as the one in [6] when the problem is reformulated as a block sparsity one. Similarly, for the same special case our definition of the NSP becomes similar to the one in [35].

We would also like to note that the hierarchy of such sparsity problems depends on their dimension. For example, a joint sparsity problem with $M = 1$ becomes the standard sparsity model. In that sense, joint sparsity models generalize standard sparsity models. The hierarchy of sparsity models is illustrated in the Venn diagram of Fig. 1.

F. Extensions

Several extensions of this formulation and the work in this paper are possible, but beyond our scope. For example, the analysis we provide is in the exactly sparse, noiseless case. As with classical compressed sensing, it is possible to accommodate sampling in the presence of noise. It is also natural to consider the extension of this work to sampling signals that are not k-sparse in a fusion frame representation but can be very well approximated by such a representation. (However, see Section IV-D.)

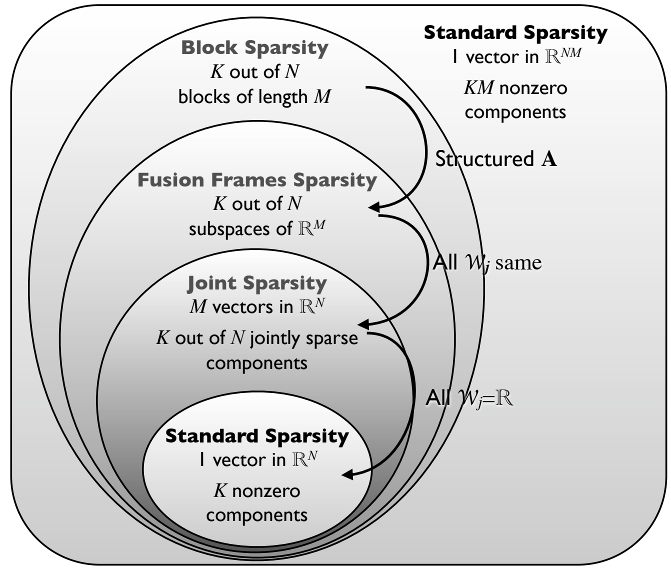

Fig. 1. Hierarchy of Sparsity Models

The richness of fusion frames also allows us to consider richer sampling matrices. Specifically, it is possible to consider sampling operators consisting of different matrices, each operating on a separate subspace of the fusion frame. Such extensions open the use of ℓ_1 methods to general vector-valued mathematical objects, to the general problem of sampling such objects [8], and to general model-based CS problems [28].

IV. Deterministic Recovery Conditions

In this section we derive conditions on $\mathbf{c}^0$ and $\mathbf{A}$ so that $\mathbf{c}^0$ is the unique solution of (P_0) as well as of (P_1). Our approach uses the generalized notions of *null space property*, *coherence* and the *restricted isometry property*, all commonly used measures of morphological difference between the vectors of a measuring matrix.

A. Fusion Null Space Property

We first prove Thm. 3.4, which demonstrates that the fusion NSP guarantees recovery, similarly to the standard CS setup [20]. This notion will also be useful later to prove recovery bounds using the fusion coherence and using the fusion restricted isometry constants.

Proof of Theorem 3.4: Assume first that the fusion NSP holds. Let $\mathbf{c}^0$ be a vector with $\|\mathbf{c}^0\|_{2,0} \le k$, and let $\mathbf{c}^1$ be an arbitrary solution of the system $\mathbf{AU}(\mathbf{c}) = \mathbf{Y}$, and set

$$\mathbf{h} = \mathbf{c}^1 - \mathbf{c}^0.$$

Letting S denote the support of $\mathbf{c}^0$, we obtain

$$\begin{aligned}\|\mathbf{c}^1\|_{2,1} - \|\mathbf{c}^0\|_{2,1} &= \|\mathbf{c}^1_S\|_{2,1} + \|\mathbf{c}^1_{S^c}\|_{2,1} - \|\mathbf{c}^0_S\|_{2,1} \\ &\ge \|\mathbf{h}_{S^c}\|_{2,1} - \|\mathbf{h}_S\|_{2,1}.\end{aligned}$$

This term is greater than zero for any $\mathbf{h} \neq 0$ provided that

$$\|\mathbf{h}_{S^c}\|_{2,1} > \|\mathbf{h}_S\|_{2,1} \tag{10}$$

or, in other words,

$$\tfrac{1}{2}\|\mathbf{h}\|_{2,1} > \|\mathbf{h}_S\|_{2,1}, \tag{11}$$

which is ensured by the fusion NSP.

Conversely, assume that all vectors $\mathbf{c}$ with $\|\mathbf{c}\|_0 \leq k$ are recovered using (P_1). Then, for any $\mathbf{h} \in \mathcal{N} \setminus \{0\}$ and any $S \subset \{1, \dots, N\}$ with $|S| \leq k$, the k-sparse vector $\mathbf{h}_S$ is the unique minimizer of $\|\mathbf{c}\|_{2,1}$ subject to $\mathbf{Ac} = \mathbf{Ah}_S$. Further, observe that $\mathbf{A}(-\mathbf{h}_{S^c}) = \mathbf{A}(\mathbf{h}_S)$ and $-\mathbf{h}_{S^c} \neq \mathbf{h}_S$, since $\mathbf{h} \in \mathcal{N} \setminus \{0\}$. Therefore, $\|\mathbf{h}_S\|_{2,1} < \|\mathbf{h}_{S^c}\|_{2,1}$, which is equivalent to the fusion NSP because $\mathbf{h}$ was arbitrary. ∎

B. Fusion Coherence

The fusion coherence is an adaptation of the coherence notion to our more complicated situation involving the *angles* between the subspaces generated by the bases $\mathbf{U}_j$, $j = 1, \dots, N$. In other words, here we face the problem of recovery of *vector-valued* (instead of scalar-valued) components and our definition is adapted to handle this.

Since the $\mathbf{P}_j$'s are projection matrices, we can also rewrite the definition of fusion coherence as

$$\mu_f = \max_{j \neq k} \left[|\langle \mathbf{a}_j, \mathbf{a}_k \rangle| \cdot |\lambda_{\max}(\mathbf{P}_j \mathbf{P}_k)|^{1/2} \right]$$

with $\lambda_{\max}$ denoting the largest eigenvalue, simply due to the fact that the eigenvalues of $\mathbf{P}_k\mathbf{P}_j\mathbf{P}_k$ and $\mathbf{P}_j\mathbf{P}_k$ coincide. Indeed, if λ is an eigenvalue of $\mathbf{P}_j\mathbf{P}_k$ with corresponding eigenvector v then $\mathbf{P}_k\mathbf{P}_j\mathbf{P}_k v = \lambda \mathbf{P}_k v$. Since $\mathbf{P}_k^2 = \mathbf{P}_k$ this implies that $\mathbf{P}_k\mathbf{P}_j\mathbf{P}_k(\mathbf{P}_k v) = \lambda \mathbf{P}_k v$ so that λ is an eigenvalue of $\mathbf{P}_k\mathbf{P}_j\mathbf{P}_k$ with eigenvector $\mathbf{P}_k v$. Let us also remark that $|\lambda_{\max}(\mathbf{P}_j\mathbf{P}_k)|^{1/2}$ equals the largest absolute value of the cosines of the principle angles between $\mathcal{W}_j$ and $\mathcal{W}_k$.

Before we continue with the proof of the recovery condition in Theorem 3.5, let us for a moment consider the following special cases of this theorem.

a) Case $M = 1$: In this case the projection matrices equal 1, and hence the problem reduces to the classical recovery problem $\mathbf{Ax} = \mathbf{y}$ with $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{y} \in \mathbb{R}^n$. Thus our result reduces to the result obtained in [36], and the fusion coherence coincides with the commonly used mutual coherence, i.e., $\mu_f = \max_{j \neq k} |\langle \mathbf{a}_j, \mathbf{a}_k \rangle|$.

b) Case $\mathcal{W}_j = \mathbb{R}^M$ for all j: In this case the problem becomes the standard joint sparsity recovery. We recover a matrix $\mathbf{X}^0 \in \mathbb{R}^{N \times M}$ with few non-zero rows from knowledge of $\mathbf{AX}^0 \in \mathbb{R}^{n \times M}$, without any constraints on the structure of each row of $\mathbf{X}^0$. The recovered matrix is the fusion frame representation of the sparse vector and each row j represents a signal in the subspace $\mathcal{W}_j$ (the general case has the constraint that $\mathbf{X}^0$ is required to be of the form $\mathbf{U}(\mathbf{c}^0)$). Again fusion coherence coincides with the commonly used mutual coherence, i.e., $\mu_f = \max_{j \neq k} |\langle \mathbf{a}_j, \mathbf{a}_k \rangle|$.

c) Case $\mathcal{W}_j \perp \mathcal{W}_k$ for all j, k: In this case the fusion coherence becomes 0. And this is also the correct answer, since in this case there exists precisely one solution of the system $\mathbf{AU}(\mathbf{c}) = \mathbf{Y}$ for a given $\mathbf{Y}$. Hence Condition (6) becomes meaningless.

d) General Case: In the general case we can consider two scenarios: either we are given the subspaces $(\mathcal{W}_j)_j$ or we are given the measuring matrix $\mathbf{A}$. In the first situation we face the task of choosing the measuring matrix such that μ_f is as small as possible. Intuitively, we would choose the vectors $(\mathbf{a}_j)_j$ so that a pair $(\mathbf{a}_j, \mathbf{a}_k)$ has a large angle if the associated two subspaces $(\mathcal{W}_j, \mathcal{W}_k)$ have a small angle, hence balancing the two factors and try to reduce the maximum. In the second situation, we can use a similar strategy now designing the subspaces $(\mathcal{W}_j)_j$ accordingly.

For the proof of Theorem 3.5 we first derive a reformulation of the equation $\mathbf{AU}(\mathbf{c}) = \mathbf{Y}$. For this, let $\mathbf{P}_j$ denote the orthogonal projection onto $\mathcal{W}_j$ for each $j = 1, \dots, N$, set $\mathbf{A_P}$ as in (4) and define the map $\varphi_k : \mathbb{R}^{k \times M} \to \mathbb{R}^{kM}$, $k \geq 1$ by

$$\varphi_k(\mathbf{Z}) = \varphi_k \begin{pmatrix} \mathbf{z}_1 \\ \hline \vdots \\ \hline \mathbf{z}_k \end{pmatrix} = (\mathbf{z}_1 \dots \mathbf{z}_k)^T,$$

i.e., the concatenation of the rows. Then it is easy to see that

$$\mathbf{AU}(\mathbf{c}) = \mathbf{Y} \quad \Leftrightarrow \quad \mathbf{A_P}\varphi_N(\mathbf{U}(\mathbf{c})) = \varphi_n(\mathbf{Y}). \tag{12}$$

We now split the proof of Theorem 3.5 into two lemmas, and wish to remark that many parts are closely inspired by the techniques employed in [36], [31]. We first show that $\mathbf{c}^0$ satisfying (6) is the *unique* solution of (P_1).

Lemma 4.1: If there exists a solution $\mathbf{U}(\mathbf{c}^0) \in \mathbb{R}^{N \times M}$ of the system $\mathbf{AU}(\mathbf{c}) = \mathbf{Y}$ with $\mathbf{c}^0$ satisfying (6), then $\mathbf{c}^0$ is the unique solution of (P_1).

Proof: We aim at showing that the condition on the fusion coherence implies the fusion NSP. To this end, let $\mathbf{h} \in \mathcal{N} \setminus \{0\}$, i.e., $\mathbf{AU}(\mathbf{h}) = 0$. By using the reformulation (12), it follows that

$$\mathbf{A_P}\varphi_N(\mathbf{U}(\mathbf{h})) = 0.$$

This implies that

$$\mathbf{A_P^*}\mathbf{A_P}\varphi_N(\mathbf{U}(\mathbf{h})) = 0.$$

Defining $\mathbf{a}_j$ by $\mathbf{a}_j = (a_{ij})_i$ for each j, the previous equality can be computed to be

$$(\langle \mathbf{a}_j, \mathbf{a}_k \rangle \mathbf{P}_j \mathbf{P}_k)_{jk} \varphi_N(\mathbf{U}(\mathbf{h})) = 0.$$

Recall that we have required the vectors $\mathbf{a}_j$ to be normalized. Hence, for each j,

$$\mathbf{U}_j \mathbf{h}_j = -\sum_{k \neq j} \langle \mathbf{a}_j, \mathbf{a}_k \rangle \mathbf{P}_j \mathbf{P}_k \mathbf{U}_k \mathbf{h}_k.$$

Since $\|\mathbf{U}_j\mathbf{h}_j\|_2 = \|\mathbf{h}_j\|_2$ for any j, this gives

$$\begin{aligned} \|\mathbf{h}_j\|_2 &\leq \sum_{k \neq j} |\langle \mathbf{a}_j, \mathbf{a}_k \rangle| \cdot \|\mathbf{P}_j \mathbf{P}_k\|_{2\to 2} \|\mathbf{h}_k\|_2 \\ &\leq \mu_f(\|\mathbf{h}\|_{2,1} - \|\mathbf{h}_j\|_2), \end{aligned}$$

which implies

$$\|\mathbf{h}_j\|_2 \leq (1 + \mu_f^{-1})^{-1} \|\mathbf{h}\|_{2,1}.$$

Thus, we have

$$\begin{aligned}\|\mathbf{h}_S\|_{2,1} &\leq \#(S)\cdot(1+\mu_f^{-1})^{-1}\|\mathbf{h}\|_{2,1}\\ &= \|\mathbf{c}^0\|_{2,0}\cdot(1+\mu_f^{-1})^{-1}\|\mathbf{h}\|_{2,1}.\end{aligned}$$

Concluding, (6) and the fusion null space property show that $\mathbf{h}$ satisfies (11) unless $\mathbf{h}=0$, which implies that $\mathbf{c}^0$ is the unique minimizer of (P_1) as claimed. ■

Using Lemma 4.1 it is easy to show the following lemma.

Lemma 4.2: If there exists a solution $\mathbf{U}(\mathbf{c}^0)\in\mathbb{R}^{N\times M}$ of the system $\mathbf{AU}(\mathbf{c})=\mathbf{Y}$ with $\mathbf{c}^0$ satisfying (6), then $\mathbf{c}^0$ is the unique solution of (P_0).

Proof: Assume $\mathbf{c}^0$ satisfies (6) and $\mathbf{AU}(\mathbf{c}^0)=\mathbf{Y}$. Then, by Lemma 4.1, it is the unique solution of (P_1). Assume there is a $\tilde{\mathbf{c}}$ satisfying $\mathbf{AU}(\tilde{\mathbf{c}})=\mathbf{Y}$ such that $\|\tilde{\mathbf{c}}\|_{2,0}\leq\|\mathbf{c}^0\|_{2,0}$. Then $\tilde{\mathbf{c}}$ also satisfies (6) and again by Lemma 4.1 $\tilde{\mathbf{c}}$ is also the unique solution to (P_1). But this means that $\tilde{\mathbf{c}}=\mathbf{c}^0$ and $\mathbf{c}^0$ is the unique solution to (P_0). ■

We observe that Theorem 3.5 now follows immediately from Lemmas 4.1 and 4.2.

C. Fusion Restricted Isometry Property

Finally, we consider the condition for sparse recovery using the restricted isometry property (RIP) of the sampling matrix. The RIP property on the sampling matrix, first introduced in [1], complements the null space propery and the mutual coherence conditions. Definition 3.3 generalizes it for the fusion frame setup. Informally, we say that $(\mathbf{A},(\mathcal{W}_j)_{j=1}^N)$ satisfies the *fusion restricted isometry property* (FRIP) if δ_k is small for reasonably large k. Note that we obtain the classical definition of the RIP of $\mathbf{A}$ if $M=1$ and all the subspaces $\mathcal{W}_j$ have dimension 1. Using this property we can prove Theorem 3.6.

Proof of Theorem 3.6: The proof proceeds analogously to the one of Theorem 2.6 in [26], that is, we establish the fusion NSP. The claim will then follow from Theorem 3.4.

Let us first note that

$$|\langle\mathbf{A_P u},\mathbf{A_P v}\rangle|\leq\delta_k\|\mathbf{u}\|_{2,2}\|\mathbf{v}\|_{2,2}$$

for all $\mathbf{u}=(\mathbf{u}_1,\ldots,\mathbf{u}_N),\mathbf{v}=(\mathbf{v}_1,\ldots,\mathbf{v}_N)\in\mathbb{R}^{MN}$, $\mathbf{u}_j,\mathbf{v}_j\in\mathbb{R}^M$, with $\operatorname{supp}\mathbf{u}=\{j:\mathbf{u}_j\neq0\}\cap\operatorname{supp}\mathbf{v}=\emptyset$ and $\|\mathbf{u}\|_{2,0}+\|\mathbf{v}\|_{2,0}\leq k$. This statement follows completely analogously to the proof of Proposition 2.5(c) in [32], see also [37], [20].

Now let $\mathbf{h}\in\mathcal{N}=\{\mathbf{h}:\mathbf{AU}(\mathbf{h})=0\}$ be given. Using the reformulation (12), it follows that

$$\mathbf{A_P}\varphi_N(\mathbf{U}(\mathbf{h}))=0.$$

In order to show the fusion NSP it is enough to consider an index set S_0 of size k of largest components $\|\mathbf{h}_j\|_2$, i.e., $\|\mathbf{h}_j\|_2\geq\|\mathbf{h}_i\|_2$ for all $j\in S_0$, $i\in S_0^c=\{1,\ldots,N\}\setminus S_0$. We partition S_0^c into index sets $S_1,S_2,\ldots$ of size k (except possibly the last one), such that S_1 is an index set of largest components in S_0^c, S_2 is an index set of largest components in $(S_0\cup S_0)^c$, etc. Let $\mathbf{h}_{S_i}$ be the vector that coincides with $\mathbf{h}$ on S_i and is set to zero outside. In view of $\mathbf{h}\in\mathcal{N}$ we have $\mathbf{AU}(\mathbf{h}_{S_0})=\mathbf{AU}(-\mathbf{h}_{S_1}-\mathbf{h}_{S_2}-\cdots)$. Now set $\mathbf{z}=\varphi_N(\mathbf{U}(\mathbf{h}))$ and $\mathbf{z}_{S_i}=\varphi_N(\mathbf{U}(\mathbf{h}_{S_i}))$. It follows that $\mathbf{A_P}(\mathbf{z}_{S_0})=\mathbf{A_P}(-\sum_{i\geq1}\mathbf{z}_{S_i})$. By definition of the FRIP we obtain

$$\begin{aligned}\|\mathbf{h}_{S_0}\|_{2,2}^2 &= \|\mathbf{z}_{S_0}\|_{2,2}^2\leq\frac{1}{1-\delta_k}\|\mathbf{A_P z}_{S_0}\|_{2,2}^2\\ &= \frac{1}{1-\delta_k}\Big\langle\mathbf{A_P z}_{S_0},\mathbf{A_P}\big(-\sum_{i\geq1}\mathbf{z}_{S_i}\big)\Big\rangle\\ &\leq \frac{1}{1-\delta_k}\sum_{i\geq1}|\langle\mathbf{A_P z}_{S_0},\mathbf{A_P}(-\mathbf{z}_{S_i})\rangle|\\ &\leq \frac{\delta_{2k}}{1-\delta_k}\|\mathbf{z}_{S_0}\|_{2,2}\cdot\sum_{i\geq1}\|\mathbf{z}_{S_i}\|_{2,2}.\end{aligned}$$

Using that $\delta_k\leq\delta_{2k}$ and dividing by $\|\mathbf{z}_{S_0}\|_{2,2}$ yields

$$\|\mathbf{z}_{S_0}\|_{2,2}\leq\frac{\delta_{2k}}{1-\delta_{2k}}\sum_{i\geq1}\|\mathbf{z}_{S_i}\|_{2,2}.$$

By construction of the sets S_i we have $\|\mathbf{z}_j\|_2\leq\frac{1}{k}\sum_{\ell\in S_{i-1}}\|\mathbf{z}_\ell\|_2$ for all $j\in S_i$, hence,

$$\|\mathbf{z}_{S_i}\|_{2,2}=\big(\sum_{j\in S_i}\|\mathbf{z}_j\|_2^2\big)^{1/2}\leq\frac{1}{\sqrt{k}}\|\mathbf{z}_{S_{i-1}}\|_{2,1}.$$

The Cauchy-Schwarz inequality yields

$$\begin{aligned}\|\mathbf{h}_{S_0}\|_{2,1} &\leq \sqrt{k}\|\mathbf{h}_{S_0}\|_{2,2}\leq\frac{\delta_{2k}}{1-\delta_{2k}}\sum_{i\geq1}\|\mathbf{z}_{S_{i-1}}\|_{2,1}\\ &\leq \frac{\delta_{2k}}{1-\delta_{2k}}(\|\mathbf{z}_{S_0}\|_{2,1}+\|\mathbf{z}_{S_0^c}\|_{2,1})<\frac{1}{2}\|\mathbf{h}\|_{2,1},\end{aligned}$$

where we used the assumption $\delta_{2k}<1/3$. Hence, the fusion null space property follows. ■

Having proved that the FRIP ensures signal recovery, our next proposition relates the classical RIP with our newly introduced FRIP. Let us note, however, that using the RIP to guarantee the FRIP does not take into account any properties of the fusion frame, so it is sub-optimal—especially if the subspaces of the fusion frame are orthogonal or almost orthogonal.

Proposition 4.3: Let $\mathbf{A}\in\mathbb{R}^{n\times N}$ with classical restricted isometry constant $\tilde{\delta}_k$, that is,

$$(1-\tilde{\delta}_k)\|y\|_2^2\leq\|\mathbf{A}y\|_2^2\leq(1+\tilde{\delta}_k)\|y\|_2^2$$

for all k-sparse $y\in\mathbb{R}^N$. Let $(\mathcal{W}_j)_{j=1}^N$ be an arbitrary fusion frame for $\mathbb{R}^M$. Then the fusion restricted isometry constant δ_k of $(\mathbf{A},(\mathcal{W})_{j=1}^N)$ satisfies $\delta_k\leq\tilde{\delta}_k$.

Proof: Let $\mathbf{c}$ satisfy $\|\mathbf{c}\|_0\leq k$, and denote the columns of the matrix $\mathbf{U}(\mathbf{c})$ by $\mathbf{u}_1,\ldots,\mathbf{u}_M$. The condition $\|\mathbf{c}\|_0\leq k$ implies that each $\mathbf{u}_i$ is k-sparse. Since $\mathbf{A}$ satisfies the RIP of order k with constant δ_k, we obtain

$$\begin{aligned}\|\mathbf{AU}(\mathbf{c})\|_{2,2}^2 &= \sum_{i=1}^M\|\mathbf{Au}_i\|_2^2\leq(1+\delta_k)\sum_{i=1}^M\|\mathbf{u}_i\|_2^2\\ &= (1+\delta_k)\|\mathbf{U}(\mathbf{c})\|_{2,2}^2=(1+\delta_k)\|\mathbf{c}\|_{2,2}^2\end{aligned}$$

as well as

$$\begin{aligned}\|\mathbf{AU}(\mathbf{c})\|_{2,2}^2 &= \sum_{i=1}^M\|\mathbf{Au}_i\|_2^2\geq(1-\delta_k)\sum_{i=1}^M\|\mathbf{u}_i\|_2^2\\ &= (1-\delta_k)\|\mathbf{U}(\mathbf{c})\|_{2,2}^2=(1-\delta_k)\|\mathbf{c}\|_{2,2}^2.\end{aligned}$$

This proves the proposition because $\mathbf{AU}(\mathbf{c})=\mathbf{A_P U}(\mathbf{c})$. ■

D. Additional Remarks and Extensions

Of course, it is possible to extend the proof of Theorem 3.6 in a similar manner to [1], [37], [38], [39] such that we can accommodate measurement noise and signals that are well approximated by sparse fusion frame representation. We state the analog of the main theorem of [37] without proof.

Theorem 4.4: Assume that the fusion restricted isometry constant δ_{2k} of $(\mathbf{A}, (\mathcal{W}_j)_{j=1}^N)$ satisfies

$$\delta_{2k} < \Delta := \sqrt{2} - 1 \approx 0.4142.$$

For $\mathbf{x} \in \mathcal{H}$, let noisy measurements $\mathbf{Y} = \mathbf{A}\mathbf{x}+\eta$ be given with $\|\eta\|_2 \leq \epsilon$. Let $\mathbf{c}^{\#}$ be the solution of the convex optimization problem

$$\min \|\mathbf{c}\|_{2,1} \quad \text{subject to} \quad \|\mathbf{A}\mathbf{U}(\mathbf{c}) - \mathbf{Y}\|_{2,2} \leq \eta,$$

and set $\mathbf{x}^{\#} = \mathbf{U}(\mathbf{c}^{\#})$. Then

$$\|\mathbf{x} - \mathbf{x}^{\#}\|_{2,2} \leq C_1 \eta + C_2 \frac{\|\mathbf{x}^k - \mathbf{x}\|_{2,1}}{\sqrt{k}}$$

where $\mathbf{x}_k$ is obtained from $\mathbf{x}$ be setting to zero all components except the k largest in norm. The constants $C_1, C_2 > 0$ only depend on δ_{2k} (or rather on $\Delta - \delta_{2k}$).

V. Probabilistic Analysis

A. General Recovery Condition

We start our analysis by deriving a recovery condition on the measurement matrix and the signal which the reader might want to compare with [40], [41]. Given a matrix $\mathbf{X} \in \mathbb{R}^{N \times M}$, we let $\mathrm{sgn}(\mathbf{X}) \in \mathbb{R}^{N \times M}$ denote the matrix which is generated from $\mathbf{X}$ by normalizing each entry X_{ji} by the norm of the corresponding row $\mathbf{X}_{j,\cdot}$. More precisely,

$$\mathrm{sgn}(\mathbf{X})_{ji} = \begin{cases} \frac{X_{ji}}{\|\mathbf{X}_{j,\cdot}\|_2} & \text{if} \quad \|\mathbf{X}_{j,\cdot}\|_2 \neq 0, \\ 0 & \text{if} \quad \|\mathbf{X}_{j,\cdot}\|_2 = 0. \end{cases}$$

Column vectors are defined similarly by $\mathbf{X}_{\cdot,i}$.

Under a certain condition on A, which is dependent on the support of the solution, we derive the result below on unique recovery. To phrase it, let $(\mathcal{W}_j)_{j=1}^N$ be a fusion frame with associated orthogonal bases $(\mathbf{U}_j)_{j=1}^N$ and orthogonal projections $(\mathbf{P}_j)_{j=1}^N$, and recall the definition of the notion $\mathbf{U}(\mathbf{c})$ in Section III. Then, for some support set $S = \{j_1, \ldots, j_{|S|}\} \subseteq \{1, \ldots, N\}$ of $\mathbf{U}(\mathbf{c})$, we let

$$\mathbf{A}_S = (\mathbf{A}_{\cdot,j_1} \cdots \mathbf{A}_{\cdot,j_{|S|}}) \in \mathbb{R}^{n \times |S|}$$

and

$$\mathbf{U}(\mathbf{c})_S = \begin{pmatrix} \mathbf{c}_{j_1}^T \mathbf{U}_{j_1}^T \\ \hline \vdots \\ \hline \mathbf{c}_{j_{|S|}}^T \mathbf{U}_{j_{|S|}}^T \end{pmatrix} \in \mathbb{R}^{|S| \times M}.$$

Before stating the theorem, we wish to remark that its proof uses similar ideas as the analog proof in [7]. We however state all details for the convenience of the reader.

Theorem 5.1: Retaining the notions from the beginning of this section, we let $\mathbf{c}_j \in \mathbb{R}^{m_j}$, $j = 1 \ldots, N$ with $S = \mathrm{supp}(\mathbf{c}) = \{j : \mathbf{c}_j \neq 0\}$. If $\mathbf{A}_S$ is non-singular and there exists a matrix $\mathbf{H} \in \mathbb{R}^{n \times M}$ such that

$$\mathbf{A}_S^T \mathbf{H} = \mathrm{sgn}(\mathbf{U}(\mathbf{c})_S) \tag{13}$$

and

$$\|\mathbf{H}^T \mathbf{A}_{\cdot,j}\|_2 < 1 \quad \text{for all } j \notin S, \tag{14}$$

then $\mathbf{U}(\mathbf{c})$ is the unique solution of (P_1).

Proof: Let $\mathbf{c}_j \in \mathbb{R}^{m_j}$, $j = 1 \ldots, N$ be a solution of $\mathbf{Y} = \mathbf{A}\mathbf{U}(\mathbf{c})$, set $S = \mathrm{supp}(\mathbf{U}(\mathbf{c}))$, and suppose $\mathbf{A}_S$ is non-singular and the hypotheses (13) and (14) are satisfied for some matrix $\mathbf{H} \in \mathbb{R}^{n \times M}$. Let $\mathbf{c}'_j \in \mathbb{R}^{m_j}$, $j = 1 \ldots, N$ with $\mathbf{c}'_{j_0} \neq \mathbf{c}_{j_0}$ for some j_0 be a different set of coefficient vectors which satisfies $\mathbf{Y} = \mathbf{A}\mathbf{U}(\mathbf{c}')$. To prove our result we aim to establish that

$$\|\mathbf{c}\|_{2,1} < \|\mathbf{c}'\|_{2,1}. \tag{15}$$

We first observe that

$$\begin{aligned} \|\mathbf{c}\|_{2,1} &= \|\mathbf{U}(\mathbf{c})\|_{2,1} = \|\mathbf{U}(\mathbf{c})_S\|_{2,1} \\ &= \mathrm{tr}\left[\mathrm{sgn}(\mathbf{U}(\mathbf{c})_S)(\mathbf{U}(\mathbf{c})_S)^T\right]. \end{aligned}$$

Set $S' = \mathrm{supp}(\mathbf{U}(\mathbf{c}'))$, apply (13), and exploit properties of the trace,

$$\begin{aligned} \|\mathbf{c}\|_{2,1} &= \mathrm{tr}\left[\mathbf{A}_S^T \mathbf{H} (\mathbf{U}(\mathbf{c})_S)^T\right] = \mathrm{tr}\left[(\mathbf{A}\mathbf{U}(\mathbf{c}))^T \mathbf{H}\right] \\ &= \mathrm{tr}\left[(\mathbf{A}\mathbf{U}(\mathbf{c}'))^T \mathbf{H}\right] \\ &= \mathrm{tr}\left[(\mathbf{A}\mathbf{U}(\mathbf{c}'_S))^T \mathbf{H}\right] + \mathrm{tr}\left[(\mathbf{A}\mathbf{U}(\mathbf{c}'_{S^c}))^T \mathbf{H}\right]. \end{aligned}$$

Now use the Cauchy-Schwarz inequality to obtain

$$\begin{aligned} \|\mathbf{c}\|_{2,1} &\leq \sum_{j \in S} \|(\mathbf{U}(\mathbf{c}')_S)_{j,\cdot}\|_2 \|(\mathbf{H}^T \mathbf{A}_S)_{\cdot,j}\|_2 \\ &\quad + \sum_{j \in S^c} \|(\mathbf{U}(\mathbf{c}')_{S^c})_{j,\cdot}\|_2 \|(\mathbf{H}^T \mathbf{A}_{S^c})_{\cdot,j}\|_2 \\ &\leq \max_{j \in S} \|(\mathbf{H}^T \mathbf{A}_S)_{\cdot,j}\|_2 \|\mathbf{c}'_S\|_{2,1} \\ &\quad + \max_{j \in S^c} \|(\mathbf{H}^T \mathbf{A}_{S^c})_{\cdot,j}\|_2 \|\mathbf{c}'_{S^c}\|_{2,1} \\ &< \|\mathbf{c}'_S\|_{2,1} + \|\mathbf{c}'_{S^c}\|_{2,1} = \|\mathbf{c}'\|_{2,1}. \end{aligned}$$

The strict inequality follows from $\|\mathbf{c}'_{S^c}\|_1 > 0$, which is true because otherwise $\mathbf{c}'$ would be supported on S. The equality $\mathbf{A}\mathbf{U}(\mathbf{c}) = \mathbf{A}\mathbf{U}(\mathbf{c}')$ would then be in contradiction to the injectivity of $\mathbf{A}_S$ (recall that $\mathbf{c} \neq \mathbf{c}'$). This concludes the proof. ■

The matrix $\mathbf{H}$ exploited in Theorem 5.1 might be chosen as

$$\mathbf{H} = (\mathbf{A}_S^\dagger)^T \mathrm{sgn}(\mathbf{U}(\mathbf{c})_S)$$

to satisfy (13). This particular choice will in fact be instrumental for the average case result we are aiming for. For now, we obtain the following result as a corollary from Theorem 5.1.

Corollary 5.2: Retaining the notions from the beginning of this section, we let $\mathbf{c}_j \in \mathbb{R}^{m_j}$, $j = 1 \ldots, N$, with $S = \mathrm{supp}(\mathbf{U}(\mathbf{c}))$. If $\mathbf{A}_S$ is non-singular and

$$\|\mathrm{sgn}(\mathbf{U}(\mathbf{c})_S)^T \mathbf{A}_S^\dagger \mathbf{A}_{\cdot,j}\|_2 < 1 \quad \text{for all } j \notin S, \tag{16}$$

then $\mathbf{U}(\mathbf{c})$ is the unique solution of (P_1).

For later use, we will introduce the matrices $\tilde{\mathbf{U}}_j \in \mathbb{R}^{M \times Nm}$ defined by

$$\tilde{\mathbf{U}}_j = (\mathbf{0}_{M \times m} | \cdots | \mathbf{0}_{M \times m} | \mathbf{U}_j | \mathbf{0}_{M \times m} | \cdots | \mathbf{0}_{M \times m}),$$

where $\mathbf{U}_j$ is the jth block. For some $\mathbf{b} = (b_1, \ldots, b_k)^T \in \mathbb{R}^k$, we can then write $\mathbf{U}(\mathbf{c})_S^T\mathbf{b}$ as

$$\sum_{\ell=1}^{k} b_\ell \tilde{\mathbf{U}}_{j_\ell}\mathbf{X} \in \mathbb{R}^M. \tag{17}$$

Using the above and the probabilistic model in Sec. III-D to prove our main probabilistic result, Theorem 3.7.

B. Probability of Sparse Recovery for Fusion Frames

The proof of our probabilistic result, Theorem 3.7, is developed in several steps. A key ingredient is a concentration of measure result: If f is a Lipschitz function on $\mathbb{R}^K$ with Lipschitz constant L, i.e., $|f(x) - f(y)| \le L\|x - y\|_2$ for all $x, y \in \mathbb{R}^K$, $\mathbf{X}$ is a K-dimensional vector of independent standard normal random variables then [42, eq. (2.35)]

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \ge u) \le 2e^{-u^2/(2L^2)} \quad \text{for all } u > 0. \tag{18}$$

Our first lemma investigates the properties of a function related to (17) that are needed to apply the above inequality.

Lemma 5.3: Let $\mathbf{b} = (b_1, \ldots, b_k)^T \in \mathbb{R}^k$ and $S = \{j_1, \ldots, j_k\} \subseteq \{1, \ldots, N\}$. Define the function f by

$$f(\mathbf{X}) = \|\sum_{\ell=1}^{k} b_\ell \tilde{\mathbf{U}}_{j_\ell}\mathbf{X}\|_2, \quad \mathbf{X} \in \mathbb{R}^{Nm}.$$

Then the following holds.

(i) f is Lipschitz with constant $\|\sum_{\ell=1}^k b_\ell \tilde{\mathbf{U}}_{j_\ell}\|_{2\to2}$.

(ii) For a standard Gaussian vector $\mathbf{X} \in \mathbb{R}^{Nm}$ we have $\mathbb{E}[f(\mathbf{X})] \le \sqrt{m}\|\mathbf{b}\|_2$.

Proof: The claim in (i) follows immediately from

$$\begin{aligned} |f(\mathbf{X}) - f(\mathbf{Y})| &= \left|\|\sum_{\ell=1}^{k} b_\ell \tilde{\mathbf{U}}_{j_\ell}\mathbf{X}\|_2 - \|\sum_{\ell=1}^{k} b_\ell \tilde{\mathbf{U}}_{j_\ell}\mathbf{Y}\|_2\right| \\ &\le \|\left(\sum_{\ell=1}^{k} b_\ell \tilde{\mathbf{U}}_{j_\ell}\right)(\mathbf{X} - \mathbf{Y})\|_2 \\ &\le \|\sum_{\ell=1}^{k} b_\ell \tilde{\mathbf{U}}_{j_\ell}\|_{2\to2}\|\mathbf{X} - \mathbf{Y}\|_2. \end{aligned}$$

It remains to prove (ii). Obviously,

$$\begin{aligned} (\mathbb{E}f(\mathbf{X}))^2 &\le \mathbb{E}[f(\mathbf{X})^2] = \mathbb{E}\left[\sum_{i=1}^{M}\left|\sum_{\ell=1}^{k} b_\ell(\tilde{\mathbf{U}}_{j_\ell}\mathbf{X})_i\right|^2\right] \\ &= \sum_{i=1}^{M}\sum_{\ell,\ell'=1}^{k} b_\ell b_{\ell'}\mathbb{E}[(\tilde{\mathbf{U}}_{j_\ell}\mathbf{X})_i(\tilde{\mathbf{U}}_{j_{\ell'}}\mathbf{X})_i]. \end{aligned}$$

Invoking the conditions on $\mathbf{X}$,

$$\begin{aligned} \mathbb{E}[f(\mathbf{X})]^2 &\le \sum_{i=1}^{M}\sum_{\ell=1}^{k} b_\ell^2\mathbb{E}[(\tilde{\mathbf{U}}_{j_\ell}\mathbf{X})_i]^2 = \sum_{\ell=1}^{k} b_\ell^2\|\tilde{\mathbf{U}}_{j_\ell}\|_F^2 \\ &= m\|\mathbf{b}\|_2^2. \end{aligned}$$

■

Next we estimate the Lipschitz constant of the function f in the previous lemma.

Lemma 5.4: Let $\mathbf{b} = (b_1, \ldots, b_k)^T \in \mathbb{R}^k$ and $S = \{j_1, \ldots, j_k\} \subseteq \{1, \ldots, N\}$. Then

$$\|\sum_{\ell=1}^{k} b_\ell \tilde{\mathbf{U}}_{j_\ell}\|_{2\to2} \le \|\mathbf{b}\|_\infty \sqrt{1 + \max_{i\in S}\sum_{j\in S, j\ne i}\lambda_{\max}(\mathbf{P}_i\mathbf{P}_j)^{1/2}}.$$

Proof: First observe that

$$\begin{aligned} \|\sum_{\ell=1}^{k} b_\ell \tilde{\mathbf{U}}_{j_\ell}\|_{2\to2} &= \|(\sum_{\ell=1}^{k} b_\ell \tilde{\mathbf{U}}_{j_\ell})^T(\sum_{\ell=1}^{k} b_\ell \tilde{\mathbf{U}}_{j_\ell})\|_{2\to2}^{1/2} \\ &= \|\sum_{\ell,\ell'=1}^{k} b_\ell b_{\ell'}\tilde{\mathbf{U}}_{j_\ell}^T\tilde{\mathbf{U}}_{j_{\ell'}}\|_{2\to2}^{1/2}. \end{aligned}$$

Since

$$\|\sum_{\ell,\ell'=1}^{k} b_\ell b_{\ell'}\tilde{\mathbf{U}}_{j_\ell}^T\tilde{\mathbf{U}}_{j_{\ell'}}\|_{2\to2} \le \|\mathbf{b}\|_\infty^2\|(\mathbf{U}_i^T\mathbf{U}_j)_{i,j\in S}\|_{2\to2},$$

it follows that

$$\|\sum_{\ell=1}^{k} b_\ell \tilde{\mathbf{U}}_{j_\ell}\|_{2\to2} \le \|\mathbf{b}\|_\infty\|(\mathbf{U}_i^T\mathbf{U}_j)_{i,j\in S}\|_{2\to2}^{1/2}. \tag{19}$$

Next,

$$\begin{aligned} \|(\mathbf{U}_i^T\mathbf{U}_j)_{i,j\in S}\|_{2\to2} &\le \max_{i\in S}\sum_{j\in S}\|\mathbf{U}_i^T\mathbf{U}_j\|_{2\to2} \qquad (20) \\ &= 1 + \max_{i\in S}\sum_{j\in S, j\ne i}\|\mathbf{U}_i^T\mathbf{U}_j\|_{2\to2}. \end{aligned}$$

By definition of the orthogonal projections $\mathbf{P}_i$,

$$\begin{aligned} \|\mathbf{U}_i^T\mathbf{U}_j\|_{2\to2} &= \|(\mathbf{U}_i^T\mathbf{U}_j)^T\mathbf{U}_i^T\mathbf{U}_j\|_{2\to2}^{1/2} \\ &= \|\mathbf{U}_i^T\mathbf{U}_j\mathbf{U}_j^T\mathbf{U}_i\|_{2\to2}^{1/2} = \|\mathbf{P}_i\mathbf{P}_j\|_{2\to2}^{1/2} \\ &= \lambda_{\max}(\mathbf{P}_i\mathbf{P}_j)^{1/2}. \end{aligned}$$

Combining with (20),

$$\|(\mathbf{U}_i^T\mathbf{U}_j)_{i,j\in S}\|_{2\to2} \le 1 + \max_{i\in S}\sum_{j\in S, j\ne i}\lambda_{\max}(\mathbf{P}_i\mathbf{P}_j)^{1/2}. \tag{21}$$

The lemma now follows from (19), (20), and (21). ■

Now we have collected all ingredients to prove our main result, Theorem 3.7

Proof: Denote $\mathbf{b}^{(j)} = (b_1^{(j)}, \ldots, b_k^{(j)})^T = \mathbf{A}_S^\dagger\mathbf{A}_{\cdot,j} \in \mathbb{R}^k$ for all $j \notin S$ and choose $\delta \in (0, 1-\alpha^2)$. By Corollary 5.2, the probability that the minimization problem (P_1) fails to recover

$\mathbf{U}(\mathbf{c})$ from $\mathbf{Y} = \mathbf{A}\mathbf{U}(\mathbf{c})$ can be estimated as

$$\begin{aligned}
& \mathbb{P}\left(\max_{j\notin S} \|\mathrm{sgn}(\mathbf{U}(\mathbf{c})_S)^T \mathbf{b}^{(j)}\|_2 > 1\right) \\
= \;& \mathbb{P}\left(\max_{j\notin S} \|\sum_{\ell=1}^{k} b_\ell^{(j)} \|U_{j_\ell}\mathbf{X}_\ell\|_2^{-1} \tilde{\mathbf{U}}_{j_\ell}\mathbf{X}\|_2 > 1\right) \\
\leq \;& \mathbb{P}\left(\max_{j\notin S} \|\sum_{\ell=1}^{k} b_\ell^{(j)} \tilde{\mathbf{U}}_{j_\ell}\mathbf{X}\|_2 > \sqrt{(1-\delta)m}\right) \\
+ \;& \mathbb{P}\left(\max_{\ell=1,\dots,k} \|U_{j_\ell}\mathbf{X}_\ell\|_2 < \sqrt{(1-\delta)m}\right) \\
\leq \;& \sum_{j\notin S} \mathbb{P}\left(\|\sum_{\ell=1}^{k} b_\ell^{(j)} \tilde{\mathbf{U}}_{j_\ell}\mathbf{X}\|_2 > \sqrt{(1-\delta)m}\right) \\
+ \;& \sum_{\ell=1}^{k} \mathbb{P}\left(\|\mathbf{X}_\ell\|_2^2 \leq (1-\delta)m\right).
\end{aligned}$$

Since $\mathbf{X}_\ell$ is a standard Gaussian vector in $\mathbb{R}^m$ [43, Corollary 3] gives

$$\mathbb{P}\left(\|\mathbf{X}_\ell\|_2^2 \leq (1-\delta)m\right) \leq \exp(-\delta^2 m/4).$$

Furthermore, the concentration inequality (18) combined with Lemmas 5.3 and Lemma 5.4 yields

$$\begin{aligned}
& \mathbb{P}\left(\|\sum_{\ell=1}^{k} b_\ell^{(j)} \tilde{\mathbf{U}}_{j_\ell}\mathbf{X}\|_2 > \sqrt{(1-\delta)m}\right) \\
= \;& \mathbb{P}\Bigg(\|\sum_{\ell=1}^{k} b_\ell^{(j)} \tilde{\mathbf{U}}_{j_\ell}\mathbf{X}\|_2 > \|b^{(j)}\|_2\sqrt{m} \\
& \qquad + (\sqrt{1-\delta} - \|b^{(j)}\|_2)\sqrt{m}\Bigg) \\
\leq \;& \exp\left(-\frac{(\sqrt{1-\delta} - \|b^{(j)}\|_2)^2 m}{2\|b^{(j)}\|_\infty^2 \theta}\right) \\
\leq \;& \exp\left(-\frac{(\sqrt{1-\delta} - \|b^{(j)}\|_2)^2 m}{2\|b^{(j)}\|_2^2 \theta}\right) \\
\leq \;& \exp\left(-\frac{(\sqrt{1-\delta} - \alpha)^2 m}{2\alpha^2 \theta}\right).
\end{aligned}$$

Combining the above estimates yields the statement of the Theorem. ■

VI. Conclusions and Discussion

The main contribution in this paper is the generalization of standard Compressed Sensing results for sparse signals to signals that have a sparse fusion frame representation. As we demonstrated, the results generalize to fusion frames in a very nice and easy to apply way, using the mixed ℓ_2/ℓ_1 norm.

A key result in our work shows that the structure in fusion frames provides additional information that can be exploited in the measurement process. Specifically, our definition of fusion coherence demonstrates the importance of prior knowledge about the signal structure. Indeed, if we know that the signal lies in subspaces with very little overlap (i.e., where $\|\mathbf{P}_j\mathbf{P}_k\|_2$ is small in Definition 3.2) we can relax the requirement on the coherence of the corresponding vectors in the sampling matrix (i.e., $|\langle \mathbf{a}_j, \mathbf{a}_k\rangle|$ in the same definition) and maintain a low fusion coherence. This behavior emerges from the inherent structure of fusion frames.

The emergence of this behavior is evident both in the guarantees provided by the fusion coherence, and in our average case analysis. Unfortunately, our analysis of this property currently has not been incorporated in a tight approach to satisfying the Fusion RIP property, as described in Section IV-C. While an extension of such analysis for the RIP guarantees is desirable, it is still an open problem.

Our average case analysis also demonstrates that as the sparsity structure of the problem becomes more intricate, the worst case analysis can become too pessimistic for many practical cases. The average case analysis provides reassurance that typical behavior is as expected; significantly better compared to the worst case. Our results corroborate and extend similar findings for the special case of joint sparsity in [7].

Acknowledgement

G. Kutyniok would like to thank Peter Casazza, David Donoho, and Ali Pezeshki for inspiring discussions on ℓ_1 minimization and fusion frames. G. Kutyniok would also like to thank the Department of Statistics at Stanford University and the Department of Mathematics at Yale University for their hospitality and support during her visits.

References

[1] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] P. G. Casazza, G. Kutyniok, and S. Li, "Fusion Frames and Distributed Processing," *Appl. Comput. Harmon. Anal.*, vol. 25, pp. 114–132, 2008.

[4] P. Boufounos, G. Kutyniok, and H. Rauhut, "Compressed sensing for fusion frames," in *Proc. SPIE, Wavelets XIII*, vol. 7446, 2009, doi:10.1117/12.826327.

[5] Y. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inform. Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[6] Y. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *Signal Processing, IEEE Transactions on*, vol. 58, no. 6, pp. 3042 –3054, 2010.

[7] Y. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inform. Theory*, vol. 56, no. 1, pp. 505–519, 2010.

[8] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

[9] M. Fornasier and H. Rauhut, "Recovery algorithms for vector valued data with joint sparsity constraints," *SIAM J. Numer. Anal.*, vol. 46, no. 2, pp. 577–613, 2008.

[10] J. A. Tropp, "Algorithms for simultaneous sparse approximation: part II: Convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.

[11] L. Peotta and P. Vandergheynst, "Matching pursuit with block incoherent dictionaries," *Signal Processing, IEEE Transactions on*, vol. 55, no. 9, pp. 4549 –4557, Sept. 2007.

[12] M. Kowalski and B. Torrésani, "Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients," *Signal, Image and Video Processing*, vol. 3, no. 3, pp. 251–264, Sept. 2009.

[13] D. Model and M. Zibulevsky, "Signal reconstruction in sensor arrays using sparse representations," *Signal Processing*, vol. 86, no. 3, pp. 624–638, 2006.

[14] D. Malioutov, "A sparse signal reconstruction perspective for source localization with sensor arrays," Master's thesis, MIT, Cambridge, MA, July 2003.

[15] A. C. Zelinski, V. K. Goyal, E. Adalsteinsson, and L. L. Wald, "Sparsity in MRI RF excitation pulse design," in *Proc. 42nd Annual Conference on Information Sciences and Systems (CISS 2008)*, March 2008, pp. 252–257.
[16] L. Daudet, "Sparse and structured decompositions of signals with the molecular matching pursuit," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 5, pp. 1808–1816, 2006.
[17] V. Cevher, R. Chellappa, and J. H. McClellan, "Vehicle Speed Estimation Using Acoustic Wave Patterns," *IEEE Trans. Signal Processing*, vol. 57, no. 1, pp. 30–47, Jan 2009.
[18] A. Veeraraghavan, D. Reddy, and R. Raskar, "Coded Strobing Photography for High Speed Periodic Events," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2009, submitted.
[19] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, pp. 227–234, 1995.
[20] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *J. Amer. Math. Soc.*, vol. 22, pp. 211–231, 2009.
[21] Y. Zhang, "A simple proof for recoverability of ℓ_1-minimization: Go over or under?" Rice University, Department of Computational and Applied Mathematics Technical Report TR05-09, August 2005, http://www.caam.rice.edu/~yzhang/reports/tr0509.pdf.
[22] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2331–2242, 2004.
[23] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
[24] H. Rauhut, "Stability results for random sampling of sparse trigonometric polynomials," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5661–5670, 2008.
[25] G. E. Pfander and H. Rauhut, "Sparsity in time-frequency representations," *J. Fourier Anal. Appl.*, vol. 16, no. 2, pp. 233–260, 2010.
[26] H. Rauhut, "Circulant and Toeplitz matrices in compressed sensing," in *Proc. SPARS '09*, Saint-Malo, France, 2009.
[27] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inform. Theory*, vol. 56, no. 1, pp. 520 –544, 2010.
[28] R. Baraniuk, V. Cevher, M. Duarte, and C. Hedge, "Model-based compressive sensing," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.
[29] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, pp. 129–159, 2001.
[30] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inform. Theory*, vol. 52, no. 1, pp. 6–18, 2006.
[31] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3320–3325, 2003.
[32] H. Rauhut, "Compressive sensing and structured random matrices," in *Theoretical Foundations and Numerical Methods for Sparse Recovery*, ser. Radon Series Comp. Appl. Math, M. Fornasier, Ed. deGruyter, 2010.
[33] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," Rice University, Depart. Electrical and Computer Engineering Technical Report TREE-0612, Nov 2006.
[34] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 655–687, 2008.
[35] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *Signal Processing, IEEE Transactions on*, vol. 57, no. 8, pp. 3075 –3085, aug. 2009.
[36] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization," *Proc. Natl. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, 2003.
[37] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci. Paris S'er. I Math.*, vol. 346, pp. 589–592, 2008.
[38] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 395–407, 2009.
[39] S. Foucart, "A note on guaranteed sparse recovery via ℓ_1-minimization," *Appl. Comput. Harmon. Anal.*, vol. 29, no. 1, pp. 97–103, July 2010.
[40] J. J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1341–1344, 2004.
[41] J. A. Tropp, "Recovery of short, complex linear combinations via l_1 minimization," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1568–1570, 2005.
[42] M. Ledoux, *The Concentration of Measure Phenomenon*. AMS, 2001.
[43] A. Barvinok, "Measure concentration," 2005, lecture notes.

Petros T. Boufounos completed his undergraduate and graduate studies at MIT. He received the S.B. degree in Economics in 2000, the S.B. and M.Eng. degrees in Electrical Engineering and Computer Science (EECS) in 2002, and the Sc.D. degree in EECS in 2006. Since January 2009 he is with Mitsubishi Electric Research Laboratories (MERL) in Cambridge, MA. He is also a visiting scholar at the Rice University Electrical and Computer Engineering department.

Between September 2006 and December 2008, Dr. Boufounos was a post-doctoral associate with the Digital Signal Processing Group at Rice University doing research in Compressive Sensing. In addition to Compressive Sensing, his immediate research interests include signal acquisition and processing, data representations, frame theory, and machine learning applied to signal processing. He is also looking into how Compressed Sensing interacts with other fields that use sensing extensively, such as robotics and mechatronics. Dr. Boufounos has received the Ernst A. Guillemin Master Thesis Award for his work on DNA sequencing and the Harold E. Hazen Award for Teaching Excellence, both from the MIT EECS department. He has also been an MIT Presidential Fellow. Dr. Boufounos is a member of the IEEE, Sigma Xi, Eta Kappa Nu, and Phi Beta Kappa.

Gitta Kutyniok received diploma degrees in mathematics and computer science in 1996 and the Dr. rer. nat. in mathematics in 2000 from the University of Paderborn. In 2001, she spent one term as a Visiting Assistant Professor at the Georgia Institute of Technology. She then held positions as a Scientific Assistant at the University of Paderborn and the University of Gießen. From Oct. 2004 to Sept. 2005, she was funded by a Research Fellowship by the DFG-German Research Foundation and spent six months at each Washington University in St. Louis and Georgia Institute of Technology. She completed her Habilitation in mathematics in 2006 and received her venia legendi. In 2006, she was awarded a Heisenberg Fellowship by the DFG. From Apr. 2007 to Sept. 2008, she was a Visiting Fellow at Princeton University, Stanford University, and Yale University. Since October 2008, she is a Professor for Applied Analysis at the University of Osnabrück. G. Kutyniok was awarded the Weierstrass Prize for outstanding teaching of the University of Paderborn in 1998, the Research Prize of the University of Paderborn in 2003, the Prize of the University of Gießen in 2006, and the von Kaven Prize by the DFG in 2007. G. Kutyniok's research interests include applied harmonic analysis, compressed sensing, frame theory, imaging sciences, and sparse approximation.

Holger Rauhut received the diploma degree in mathematics from the Technical University of Munich in 2001. He was a member of the graduate program Applied Algorithmic Mathematics at the Technical University of Munich from 2002 until 2004, and received the Dr. rer. nat. in mathematics in 2004. After his time as a PostDoc in Wrocław, Poland in 2005 (3 months), he has been with the Numerical Harmonic Analysis Group at the Faculty of Mathematics of the University of Vienna, where he completed has habilitation degree in 2008. From March 2006 until February 2008 he was funded by an Individual Marie Curie Fellowship from the European Union. From March 2008, he is professor (Bonn Junior Fellow) at the Hausdorff Center for Mathematics and Institute for Numerical Simulation at the University of Bonn. In 2010 he was awarded a Starting Grant from the European Research Council. H. Rauhut's research interests include compressed sensing, sparse approximation, time-frequency and wavelet analysis.